Forecasting low volume products

Many retail stores have a significant number of low-volume products. These are products where sales do not occur in every week. In fact, some of these products may go many weeks between sales. For this reason, a forecasting system designed to be used at the store level must be able to forecast these types of products.

There are two different approaches to forecasting these types of products.

1. Create weekly forecasts which are very small numbers.
2. Forecast in integer values, spaced over time.

Small weekly forecasts

An example of this type of forecasting would be a product that sells at a rate of 10 per year or about 0.2 per week. This type of product can be forecast using weekly forecasts with a quantity of 0.2. (For the sake of simplicity, this example assumes a uniform sales rate over the year.)

The difficulty with this approach is not so much the forecast as the effect on replenishment planning. Whenever the projected on-hand balance goes below the safety stock, the replenishment logic will create a planned shipment. For example a safety stock (display quantity) of 2 and an on-hand balance of 2 is a fairly typical situation for many products (one to go and one to show). However, an on-hand balance of 2 and weekly forecast of 0.2 gives a projected on-hand balance of 1.98 at the end of the first week. This is below the safety stock of 2 and consequently the replenishment logic will create a planned shipment in the first week.

This planned shipment is typically within the lead time, and so it should be released and shipped to the store sometime this week.

Yet, this is not what most retailers would want done.

In this example the forecast for the week is 0.2 units, so eight times out of ten, the product will not sell during that week. Sending one or more units of the product to the store in anticipation of this low a forecast is not what most retailers would choose to do.

For many retailers, it is typical to have a large number of products in this situation because they ship small quantities of low-volume products to the stores resulting in on-hand balances that are not much greater than the safety stock (display quantity). As a result, there can be a large number of planned orders in the first few weeks, but these planned orders would typically not be shipped to the stores in anticipation of such a low probability of selling a unit during this week.

This bunching up of planned shipments in the first few weeks violates the basic rule of these systems: a valid simulation of reality. This large number of planned orders in the first time period creates a false demand on the distribution center for use in planning replenishment at the DC (because planned shipments at the stores are accumulated to become the projected demand on the store's supplier). Additionally, the large number of planned orders in the first time period also creates false demands in capacity planning, load building, and financial planning. See the explanations in the white paper for more information on dependent demand (planned shipments to stores replacing the forecast at a distribution center or supplier), financial planning, capacity planning, and transportation planning.

Forecast in integer values spaced over time

Using the example above (a product that sells at a rate of 10 per year or about 0.2 per week), this type of forecasting would represent the forecast as a quantity of one every 35 or 36 days. This will appear as one or more spikes on the forecasting graph (see the question on the forecast as a series of spikes for an illustration).

The forecast is actually stored as a quantity and a time period, for example, a quantity of 2.5 over a 90 day time period. When the forecast is read from the database, it is converted from a quantity and a time period (2.5 and 90 days) to a series of integers spaced 35 or so days apart.

This has the advantage of being more like what will actually happen since 0.2 will not sell every week, but one unit will sell every so often.

This approach also has the advantage of not bunching up all the planned orders in the first few time periods and therefore provides a more valid simulation of reality.

The key to this is a randomization calculation intended to simulate the what happens in the real world.

A forecast is a probability. A forecast of 1 every thirty-five or so days for example means that it's likely the product will sell one somewhere in the thirty-five day forecast period. For some stores, the product will sell early in the forecast period, such as day one or two for example. For other stores, the product will sell late in the forecast period, such as day 33 or 34 for example, and so on for the other days in the forecast period. There is no way to tell which stores will sell early in the forecast period and which stores will sell late in the forecast period. However, it is important to spread these forecasts out in such a way that the totals provide an even distribution over time, and are not bunched up at the beginning, middle or end of the forecast period (remember that this example has a flat or uniform selling pattern over the year.)

So, the forecasting logic attempts to duplicate what most planners and analysts know; that a unit will sell every so often. This logic makes the best possible guess at when this might be. The logic also does this in such a way that the total of these integer forecasts is an accurate a prediction of the total demand as a forecasting system can produce.

Because these forecasts are in integer quantities and are spaced over time, the planned shipments are not bunched up in the first few weeks as in the situation explained earlier. If a store has an on-hand balance of 2, a safety stock (display quantity) of 2, and a forecast for one 25 days in the future, the first planned shipment will be created for 25 days in the future. If another store has an on-hand balance of 2, a safety stock (display quantity) of 2, and a forecast for one 10 days in the future, the first planned shipment for the is product will be created for 10 days in the future, and so on for the other stores. This provides a series of planned shipments that can be used for financial planning, capacity planning, and transportation planning because they represent, in total, what is likely to happen.

What will actually happen is that a store with a forecast of one 25 days in the future may make a sale early in the forecast period, and a store with a forecast of one 10 days in the future may not make a sale until late in the period. When a store makes a sale, the on-hand balance drops below the safety stock, and a planned shipment is created for the first available shipping date. When this shipment is received at the store, the on-hand balance now equals (or exceeds the safety stock), and the process of randomizing the predictions of when a sale will occur starts over again.

In this example, we have assumed a flat or uniform selling pattern over the year. While this makes the logic easier to explain, it does not represent most products. Therefore, the logic that attempts to predict when a sale will happen also takes into account the selling pattern for the product (see profiles and weekly percentages for more on this subject). The result is that the integer forecasts of one will be more closely spaced during a period when the product sells well, and will be spaced further apart during a period when the product sells less well. For example, if the product is expected to sell once every 35 or so days on average, there may be periods during the year where the forecasts for one are spaced only 20 days apart, and also periods during the year where the forecasts for one are spaced 40 days apart. If you look closely at the illustration in the question on a forecast shown as a series of spikes, you will see that the spacing between spikes is different reflecting a selling pattern that is higher in some portions of the year.

As an aside, you may notice that the product record contains an internal field named randomizer. This is used to do the randomization calculations on these low-volume forecasts. The randomizer is a number between zero and one, and is calculated internally by the system if the product ever needs a low-volume forecast.

There are two related subjects that are good to understand when dealing with low-volume products. These are: profiles, weekly percentages, and forecasting in different sized time periods.